The ACE 2004 Evaluation Plan

Evaluation of the Recognition of ACE *Entities*, ACE *Relations* and ACE *Events*

1 INTRODUCTION

The objective of the ACE program is to develop automatic content extraction technology to support the automatic processing of source language data. This down-stream processing includes classification, filtering, and selection based on the content of the source data, i.e., based on the meaning conveyed by the language. Thus, the ACE program is dedicated to the development of technologies that automatically infer meaning from language data.

2 TASK DEFINITIONS

There are four primary ACE recognition tasks – the recognition of entities, relations, events and time expressions. These tasks require systems to process language data in documents and then to output, for each of these documents, information about the entities, relations, event and times discussed in them. This section provides an overview of the ACE tasks. For a complete description refer to the ACE annotation guidelines.[1] The form of the output that is required is defined by an XML format call "APF" and is described in appendix B.

2.1 ENTITY DETECTION AND RECOGNITION[2]

The ACE Entity Detection and Recognition task (EDR) requires that certain specified types of entities that are mentioned in the source language data be detected and that selected information about these entities be recognized and merged into a unified representation for each detected entity.

2.1.1 ENTITIES

Entity output is required for each document in which the entity is mentioned. This output includes information about the attributes of the entity and about the mentions of the entity. Entity attributes are currently limited to the *name*(s) used to refer to the entity, the entity *type*, the entity *subtype*, and the entity *class*.

The allowable ACE entity types and subtypes are listed in Table 1. Entities may have only one type and one subtype. Entity types, subtypes and classes are described in detail in the annotation guidelines. Of the classes discussed in the guidelines, output is required only for entities of class specific.

It often happens that different entities may be referred to by the same name. Despite this metonymic connection, however, such entities are regarded as separate and distinct for the purposes of the ACE program. For example, in the sentence "*Miami is growing rapidly*", Miami is a mention of a GPE entity named "Miami", whereas in the sentence "*Miami defeated Atlanta 28 to 3*" Miami is a metonymic mention of an organization entity named "Dolphins" and is distinct from the Miami GPE entity.

There are no limits on the use of inference and world knowledge in determining either the entity type or the entity attributes. The determination should represent the system's best judgment of the source's intention (i.e., the intention of the author or speaker).

Table 1 ACE Entity Types and Subtypes

Type	Subtype
Person (PER)	(none)
Organization (ORG)	Government, Commercial, Educational, Non-Profit, Other
Location (LOC)	Address, Boundary, Celestial, Land-Region-Natural, Region-Local, Region-Subnational, Region-National, Region-International, Water-Body, Other
Geo-Political Entity (GPE)[3]	Continent, Nation, State-or-Province, County-or-District, Population-Center, Other
Facility (FAC)	Building, Subarea-Building, Bounded-Area, Conduit, Path, Barrier, Plant, Other
Vehicle (VEH)	Land, Air, Water, Subarea-Vehicle, Other
Weapon (WEA)	Blunt, Exploding, Sharp, Chemical, Biological, Shooting, Projectile, Nuclear, Other

2.1.2 ENTITY MENTIONS

All mentions of each ACE entity are to be detected and output along with the entity attributes. The output for each entity mention includes the mention *type*, the location of its *head* and its *extent*, and optionally the mode of *reference* and the *role* of the entity in the context of the mention. These are described in detail in the annotation guidelines. The allowable mention types are listed in Table 2. The style of referring to an entity may assume one of two values, namely either literal or metonymic, depending on whether the reference is metonymic or not. The *reference* attribute is needed only for metonymic references, in which case it should be assigned the value metonymic.

---

[1] http://www.ldc.upenn.edu/Projects/ACE/Annotation/

[2] This is a new appellation for the Entity Detection and Tracking (EDT) task, one that more accurately represents the nature of the task and that integrates better with the other two ACE tasks: Relation Detection and Recognition (RDR) and Event Detection and Recognition (VDR).

[3] Geo-Political Entities deserve a little supplemental explanation and historical background. Originally, GPE's were not part of the ACE entity inventory. However, during the initial annotation exercises, it became clear that the same word would often imply different entity types – sometimes *location* (as in "the riots in Miami"), sometimes *organization* (as in "Miami imposed a curfew"), sometimes as *person* (as in "Miami railed against the curfew"). Even more troublesome, co-reference was sometimes observed between different underlying entity types (as in "Miami imposed a curfew because of its riots"). These issues gave rise to the definition of the hybrid Geo-Political entity type. This type can be viewed as somewhat synthetic and ad hoc, but there is also support for its conceptual reality, for example by the use of co-reference in joining different entity types.

Table 2 Scorable ACE Mention Types

Type	Description
Name (NAM)	A proper name reference to the entity
Nominal (NOM)	A common noun reference to the entity
Pronominal (PRO)	A pronoun reference to the entity
Premodifier (PRE)	A premodifier reference to the entity

2.2 RELATION DETECTION AND RECOGNITION

The ACE Relation Detection and Recognition task (RDR) requires that certain specified types of relations that are mentioned in the source language data be detected and that selected information about these relations be recognized and merged into a unified representation for each detected relation.

2.2.1 RELATIONS

An ACE relation is a relation between two ACE entities, which are called the relation arguments. Some relations are symmetric, meaning that the ordering of the two entities does not matter (e.g., "partner"). But for asymmetric relations the order does matter (e.g., "subsidiary") and for these relations the entity arguments must be assigned the correct argument number.

Relation output is required for each document in which the relation is mentioned. This output includes information about the attributes of the relation, the relation arguments, and the relation mentions. Relation attributes are the relation *type*, *subtype*, and *class*. The allowable ACE relation types and subtypes are listed in Table 3. Relations may have only one type and one subtype

Table 3 ACE Relation Types and Subtypes
(Relations marked with an * are symmetric relations.)

Type	Subtype
Physical (PHYS)	Located, Near*, Part-whole
Personal / Social (PER-SOC)	Business*, Family*, Other*
Employment / Membership / Subsidiary (EMP-ORG)	Employ-Executive, Employ-Staff, Employ-Undetermined, Member-of-Group, Partner*, Subsidiary, Other*
Agent-Artifact (ART)	User-or-Owner, Inventor-or-Manufacturer, Other
PER/ORG Affiliation (OTHER-AFF)	Ethnic, Ideology, Other
GPE Affiliation (GPE-AFF)	Citizen-or-Resident, Based-in, Other
Discourse (DISC)	(none)

The *class* of output relations is now restricted to explicit only.[4] Relation types and subtypes are described in detail in the annotation guidelines.

2.2.2 RELATION ARGUMENTS

Relation arguments are identified by their associated entity ID's and by an argument order (ARGNUM) value, either "1" or "2", indicating which of the two roles the argument plays in the relation. Allowable arguments and argument types are described in detail in the annotation guidelines.

2.3 EVENT DETECTION AND RECOGNITION

The ACE Event Detection and Recognition task (VDR) requires that certain specified types of events that are mentioned in the source language data be detected and that selected information about these events be recognized and merged into a unified representation for each detected event.

2.3.1 EVENTS

Currently there are five ACE event types, which are listed in Table 4. These ACE events are events in which ACE entities participate and play certain roles.

Table 4 ACE Event Types

Types:
Destruction/Damage (BRK)
Creation/Improvement (MAK)
Transfer of Possession or Control (GIV)
Movement (MOV)
Interaction of Agents (INT)

The ACE entities that are involved in the event are called event participants, and each participant is characterized by a role. The participant roles are listed in Table 5. The same entity may play more than one role and therefore may appear as a participant more than once.

Table 5 ACE Event Participant Roles

Role	Description
Agent	The cause of the event
Object	The entity acted upon by the event
Source	The original location (for MOV or GIV only)
Target	The resultant location (for MOV or GIV only)
Time	The time of the event
Location	The location of the event
Other	Other event participants

2.4 TIMEX2 DETECTION AND RECOGNITION

The ACE Timex2 Detection and Recognition task (TDR, a.k.a. "TERN", for Time Expression Recognition and Normalization) requires that certain specified types of time expressions (timex2) that are mentioned in the source language data be detected and that selected information about these timex2 expressions be recognized and merged into a unified representation for each detected timex2. A complete description of the current TDR task for 2004, Timex2, is available at URL http://timex2.mitre.org/. Further supplementary information regarding the evaluation is available at URL http://www.nist.gov/speech/tests/ace/ace04/.

---

[4] Previously both explicit and implicit relations were included in the scope of RDR, but implicit relations were found to be not very well defined or bounded and difficult to annotate reliably.

3 EVALUATION

Evaluation of ACE system performance will be supported for a total of six tasks in three languages. In addition, three types of sources and three processing modes are supported, as listed in Table 6.

Participation is allowed on any one or more of the six tasks and in any one or more of the three languages. For each task/language combination chosen, all source material must be processed by the system being evaluated, including all of the different source types contained in the evaluation data.

For the 2004 evaluation, the evaluation test corpus will not support the event detection tasks, and newspaper data will not be included. Also, cross-document annotation will not be available. 2004 evaluation support is summarized in Table 6.

Performance on each of the different ACE tasks is measured separately. However, since the arguments of relations and the participants in events are ACE entities, a system's performance on relations and events is strongly affected by the system's underlying performance on entities.

3.1 EVALUATION METHOD

System performance on each of the several tasks is scored using a model of the application value of system output. This overall value is the sum of the value for each system output entity (or relation or event), accumulated over all system outputs. The value of a system output is computed by comparing its attributes and associated information with the attributes and associated information of the reference entity (or relation or event) that corresponds to it. When system output information differs from that of the reference, value is lost. And when system output is spurious (i.e., there is no corresponding reference), negative value typically results. Perfect system output performance is achieved when the system output matches the reference without error. The overall score of a system is computed as the system output information relative to this perfect output:

$$System\_Value = \frac{\sum_i Value\left(sys\_output_i, reference_{map(i)}\right)}{\sum_m Value\left(reference_m, reference_m\right)} \qquad (1)$$

where *map*(*i*) selects the reference corresponding to sys_output *i* (or null if there is no corresponding reference).

Determining *map*(*i*) is a major step in evaluation because there is no immediate way of determining a correct or optimum mapping. This is especially true for EDR/RDR/VDR because the output is an abstraction that is inferred from the input data and for which there is no direct one-to-one connection to the correct (reference) output. Therefore the mapping operation is performed by first comparing each system output with each reference and then finding the mapping that produces the best (greatest) *System_Value*.

Historically, it has been found that loss of value is attributable mostly to misses (where a reference has no corresponding system output) and false alarms (where a system output has no corresponding reference). To a lesser extent, value is lost due to errors in determining attributes and other associated information in those cases where the system output actually does have a corresponding reference.

Table 6 ACE evaluation support and 2004 scope

		2004 Evaluation
Tasks:		
Entity Detection and Recognition (EDR)		✓
Entity Mention Detection (EMD)		✓
EDR Co-Reference (given correct mentions)		✓
Relation Detection and Recognition (RDR)		✓
Relation Mention Detection (RMD)		✓
RDR given correct entities		✓
Event Detection and Recognition (VDR)		
Event Mention Detection (VMD)		
Timex2 Detection and Recognition (TDR)		✓
Languages:		
English		✓
Chinese		✓
Arabic		✓
Mode:		
Document-Specific		✓
Cross-Document		
Database Reconciled		
Sources:		
Newswire		✓
Broadcast News	Ground truth	✓
	STT Output	✓[5]
Newspaper	Ground Truth	
	OCR Output	

3.2 EVALUATION TASKS

3.2.1 ENTITY DETECTION (EDR)

The EDR task is to infer ACE-defined entities from the source language and to recognize and output selected entity attributes and information about these entities, including information about their mentions. A major part of this problem is to identify and disambiguate entity mentions – i.e., the co-reference problem. To help focus on this essential aspect of the EDR task, the performance measure used to score EDR performance has been changed to reflect more directly a system's ability to correctly identify and collect all mentions of each entity. The *Value* formula for EDR (for a single entity) is now defined as the product of an inherent entity value and the accrued value of the mentions[6]:

---

[5] STT output will be provided for evaluation on English broadcast news only. For Arabic and Chinese, only the ground truth (manually transcribed) versions will be provided. The STT output will be provided in "CTM" format. For information about this format, please refer Appendix C.

$$Value_{sys\_entity} = Entity\_Value(sys\_entity) \cdot \sum Mention\_Value(sys\_mention_m) \tag{2}$$

where the *Entity_Value* is a function of entity type and class and the *Mention_Value* is a function of mention type. Refer to appendix A for a complete description of the EDR *Value* formula.

3.2.2 ENTITY MENTION DETECTION (EMD)

The EMD task complements the EDR task by disregarding the co-reference issue. In essence, the entity mentions are treated as though each mention were the mention of a unique entity that has only a single mention and is thus distinct and separate from all other entity mentions. This treatment allows the entity mentions to be evaluated directly, as entities, using the *Value* formula for EDR. Thus the mechanics of EMD scoring are identical to those for EDR.

3.2.3 EDR CO-REFERENCE TASK

The 2004 ACE evaluation will support an EDR co-reference evaluation by providing sites with perfect mentions for use in EDR. In order to participate in this task sites must first submit their EDR results. To accommodate a co-reference evaluation without jeopardizing the integrity of the EDR task, the ground truth mentions will not be distributed until after the EDR results are due (see section 3.6 Schedule).

The ground truth mentions will be provided in a standoff file (1 per document). Only a mention's extent and head will be provided.

3.2.4 RELATION DETECTION (RDR)

The RDR task is to infer ACE-defined relations from the source language and to recognize and output selected attributes and information about these relations, including information about their mentions. A major part of correctly recognizing relations is correctly recognizing the arguments (entities) that are related by the relation. Therefore good EDR performance is important to achieving good RDR performance. The *Value* formula for RDR (for a single relation) is defined as the product of an inherent relation value and the sum of the values of the two entities that are the relation arguments:

$$Value_{sys\_relation} = Relation\_Value(sys\_relation) \cdot \sum_a Argument\_Value(sys\_\mathrm{arg}ument_a) \tag{3}$$

where the *Relation_Value* is a function of relation type and the *Argument_Value* is the value of the entity argument as computed for EDR scoring. Refer to appendix A for a complete description of the RDR *Value* formula.

3.2.5 RELATION MENTION DETECTION (RMD)

As with entities, the RMD task complements the RDR task by disregarding the co-reference issue. In essence, the relation mentions are treated as though each mention were the mention of a unique relation that has only a single mention and is thus distinct and separate from all other relation mentions. This treatment allows the relation mentions to be evaluated directly, as relations, using the *Value* formula for RDR. Thus the mechanics of RMD scoring are identical to those for RDR. There is, however, a significant difference in computing the contribution of the relation arguments to the value of a relation mention. This is namely that the contribution of an argument is computed only for the single entity mention that is referenced in the relation mention output.

3.2.6 RDR GIVEN CORRECT ENTITIES

The 2004 ACE evaluation will support the evaluation of RDR given perfect entities. In order to participate in this task sites must first submit their RDR results. To accommodate a "perfect entity RDR" task without jeopardizing the integrity of the RDR task or the EDR co-reference task, the ground truth entities will not be distributed until after both the RDR results and the EDR co-reference results are due (see section 3.6 Schedule).

The ground truth entities will be provided in a standoff file (1 per document). They will be the ACE evaluation answer keys with all Relation information removed.

3.2.7 EVENT DETECTION (VDR)

The VDR task is to infer ACE-defined events from the source language and to recognize and output selected attributes and information about these events, including information about their mentions. A major part of correctly recognizing events is correctly recognizing the participants (entities) that participate in the event. Therefore good EDR performance is important to achieving good VDR performance. The *Value* formula for VDR (for a single event) is defined as the product of an inherent event value and the sum of the values of the entities that participate in the event:

$$Value_{sys\_event} = Event\_Value(sys\_event) \cdot \sum_p Participant\_Value(sys\_participant_p) \tag{4}$$

where the *Event_Value* is a function of event type and the *Participant_Value* is the value of the entity participant as computed for EDR scoring. Refer to appendix A for a complete description of the VDR *Value* formula.

3.2.8 EVENT MENTION DETECTION (VMD)

As with relations, the VMD task complements the VDR task by disregarding the co-reference issue. In essence, the event mentions are treated as though each mention were the mention of a unique event that has only a single mention and is thus distinct and separate from all other event mentions. This treatment allows the event mentions to be evaluated directly, as events, using the *Value* formula for VDR. Thus the mechanics of VMD scoring are identical to those for VDR. There is, however, a significant difference in computing the contribution of the event participants to the value of an event mention. This is namely that the contribution of event participants is computed only for those entities and only for the single mention of those entities that are referenced in the event mention output.

---

[6] Previously the *Value* formula did not include the accumulation of mention value over all mentions of the entity. This change highlights the importance of correctly detecting and referencing the mentions of the entities.

3.2.9 TIMEX2 DETECTION (TERN/TDR)

The evaluation of the TERN task is specified in the TERN evaluation plan.[7] In addition, NIST provides a value-based TDR evaluation. The *Value* formula for TDR (for a single timex2) is now defined as the product of the timex2 value and a value de-weighting factor for each timex2 attribute error.

$$\begin{aligned} Value_{sys\_timex} = {} & Timex\_Value(sys\_timex) \\ & \cdot \prod_i W_{Terr}(sys\_attribute_i) \end{aligned} \qquad (5)$$

Refer to appendix A for a complete description of the TDR *Value* formula.

3.3 CORPUS SUPPORT

Source language data is being provided to support research (with training and development test corpora) and evaluation (with an evaluation test corpus). ACE corpora are assembled from a variety of sources selected from broadcast news programs, newspapers, and newswire reports.

3.3.1 THE ACE 2004 TRAINING CORPUS

The Linguistic Data Consortium has newly annotated ACE training data available[8] for system development. The data is taken from broadcast news and newswire sources and is available in Arabic, Chinese and English.

Four versions of each document are provided:

- APF files (.apf.xml): The ACE Program Format has undergone a number of changes. The training corpus contains documentation describing these changes.
- ALF files (.alf.xml): The ACE LDC Format is an intermediate format similar to APF designed to store all annotation contents represented in the AG files (below).
- AG files (-pp.ag.xml): The LDC Annotation Graph Format (post-processed). LDC's internal annotation files format for ACE. These files can be viewed with LDC's annotation tool.
- Source text files (sgm and sgm.utf8): Source text files. All files, including the Chinese files, are in UTF-8. Only text between the begin text tag <TEXT> and end text tag </TEXT> are to be evaluated.

To verify data format integrity, two DTD's are distributed with the ACE training corpus. One DTD is used to verify the APF format, and one to verify the ALF format.

The ACE training corpus used data sources published between October and December 2000. Corpus statistics are in Table 7:

Table 7: 2004 ACE system training corpus statistics for release LDC2004E17. This will be an incremental release. Numbers shown represents final release on July 5, 2004.

English Resources	
Broadcast News	~57,500 words
Newswire	~57,500 words
Other	~35,000 words
Arabic Resources	
Broadcast News	~62,500 words
Newswire	~62,500 words
Other	~25,000 words
Chinese Resources	
Broadcast News	~62,000 words
Newswire	~62,500 words
Other	~25,000 words

3.3.2 THE 2004 EVALUATION CORPUS

A new evaluation data set is defined for the 2004 evaluation. The Broadcast News and Newswire data has been selected from sources originally published during the month of January 2001. Table 8 lists the statistics of the ACE 2004 evaluation corpus.

A key part of system output is the specification of entity mentions in terms of word locations in the source text. Word/phrase location information is in terms of the indices of the first and last characters of the word/phrase. EDR/RDR systems must compute these indices from the source data. Indices start with index 0 being assigned to the first character of a document. Ancillary information and annotation, which is provided as bracketed SGML tags, is not included in this count. Only characters outside of angle-bracketed expressions contribute to the character count. Also, each new line (nl or cr/lf) counts as one character.

Table 8: 2004 ACE evaluation corpus statistics.

	Arabic	English	Chinese
Broadcast News	25K words	25K words	25K words
Newswire	25K words	25K words	25K words

3.3.3 2004 EVALUATION AND SCORING CONDITIONS

Scoring will be done at the document level. This means that each entity or relation will contribute to the score for each document that mentions that entity or relation. For example, if an entity is mentioned in N different documents, that entity will contribute to the score N times.

Only one of the three processing modes will be supported for the 2004 evaluation:

- **Document-level processing.** For this processing mode, each document is processed independently of other documents.

---

[7] http://timex2.mitre.org/tern_evalplan-2004.29apr04.pdf

[8] Registered participants should contact the LDC to obtain the ACE 2004 training corpus v1.0 (LDC2004E17). Sites should inquire if an updated release is available.

No reconciliation is required (or allowed) of entities, relations or events, either across documents or with respect to a database. Thus all entities and relations mentioned in a single document must be uniquely associated and identified with that document. This means, by way of example, that if a specific person, say the US president George W. Bush, is mentioned in more than one document, then he must be represented by multiple entities – a different entity (with a globally unique ID) for each document in which he is mentioned.

There are three different source conditions:

- Newswire
- Broadcast News – ground truth
- Broadcast News – automatically produced transcripts from a speech-to-text engine.

The overall score for a system will be determined separately for each source condition. In addition, an overall system score will be computed by summing the value of all system outputs over both Newswire and Broadcast News (ground truth) and normalizing by the reference value, using equation 1. (Weighting to equalize the contribution from each source type is not anticipated. Note however that the contribution from each source type should be comparable since the amount of data from each source will be approximately equal.)

3.4 TOOLS

3.4.1 XML VALIDATION TOOLS

A java implementation of an XML validator is available from the NIST ACE web site. The XML validator will verify that a system output file conforms to the ACE DTD.[9]

3.4.2 ACE EVALUATION SOFTWARE

The ACE evaluation software is available for download from the NIST ACE web site.[10] This tool scores both EDR and RDR output.

3.5 RULES

- No changes to the system are allowed once the evaluation data are released. Adaptive systems may of course change themselves in response to the source data that they process, but…
- No human intervention is allowed prior to the submission of your test site's results to NIST.[11] This means that, in addition to disallowing modifications to your system, there must also be no modifications to or human examination of the test data.
- For each evaluation combination of task, language, and processing mode for which system output is submitted, all documents from all sources for that evaluation combination must be processed.

3.6 SCHEDULE

Date (2004)	Event
April 1st	1st release of training data. This is a partial release and will include some EDR and RDR annotations for English, Arabic & Chinese.
July 5th	Final release of training data.
July 15th	Last day to register for participation in the evaluation.[12]
August 2nd	Evaluation period begins. Test data available to be sent to sites via e-mail. Sites may request the test data at anytime between the 2nd and the 13th, but sites must return results within 24 hours of receiving the data.
August 13th	Evaluation period ends for Main ACE evaluation tasks Last day to submit results to NIST.
August 16th Noon EDT	Ground truth mentions available for EDR co-reference task
August 18th 10am EDT	EDR co-reference results due
August 18th Noon	Ground truth entities available for RDR with perfect entities task
August 20th	RDR with perfect entities results are due
August 23rd	NIST release results for Main ACE evaluation tasks
August 27th	NIST releases results for all tasks submitted before their appropriate deadline
September 20-22	Workshop in Alexandria VA, The Hilton Mark Center
September 23	TERN workshop[13]

3.7 PARTICIPATION IN ACE 2004

The NIST 2004 ACE evaluation is open to all who wish to fully participate. To officially register for the ACE evaluation, sites are required to complete the ACE Registration form located at: ftp://jaguar.ncsl.nist.gov/ace/doc/RegistrationForm-ACE_0.pdf

---

[9] The DTD's used for the ACE program, can be found at: http://www.nist.gov/speech/tests/ace/DTD

[10] The ACE evaluation tools may be accessed from the NIST ACE URL http://www.nist.gov/speech/tests/ace/phase2b/resource/

[11] It sometimes happens that a system bug is discovered during the course of processing the test data. In such a case, please consult with NIST (Mark Przybocki, phone 1-301-975-3347, or email mark.Przybocki@nist.gov) for advice. He will advise you on how to proceed. Repairs may be possible that allow a more accurate assessment of the underlying performance of a system. If this happens, modified results may be accepted, provided that an explanation of the modification is provided and provided that the original results are also submitted and documented.

[12] To register for participation in the NIST 2004 ACE evaluation, simply complete and return the registration form: ftp://jaguar.ncsl.nist.gov/ace/doc/ace_registration_2004.pdf

[13] Information regarding the Time Expression Recognition and Normalization (TERN) Evaluation may be found at http://timex2.mitre.org/tern.html

The evaluation period for the tasks EDR, EMD, RDR and RMD begins on August 2nd, 2004. Sites must request[14] the evaluation data at any time between the 2nd and the 13th, but result must be returned within 24 hours. All results for these four tasks are due by August 13th.

The tasks of EDR co-reference and RDR given correct entities requires that partial answer information be distributed. The schedule for these tasks is as defined in section 3.6.

3.8 SUBMISSION OF SYSTEM OUTPUT TO NIST

Due to the short period of time between the date system output files are due at NIST and the beginning of the workshop, it will expedite scoring and releasing of results if participants follow the outlined procedure for submitting results. This will enable quick unpacking and scoring of several site submission files with minimum human intervention.

3.8.1 PACKAGING YOUR SYSTEM OUTPUT

STEP1: Create a top level directory for each of the ***languages*** attempted:

Example: `$> mkdir chinese english`

STEP2: Create a subdirectory identifying the ***tasks*** attempted:

Example: `$> mkdir english/edr english/rdr chinese/edr`

STEP3: In each of these subdirectories make one directory for each system submitted (choose a name that identifies your site, BBN, SHEF, SRI…):

Example: `$> mkdir english/edr/NIST1_primary`

Example: `$> mkdir english/edr/NIST2_contrastive1`

Example: `$> mkdir english/rdr/NIST1_primary`

Example: `$> mkdir chinese/edr/NIST1_primary`

STEP4: Deposit all system output files in the appropriate system directory. Include a system description in this same directory (see section 8.2 for details).

STEP5: Create a compressed tar file of your results and transfer them to NIST by FTP (ftp://ijaguar.ncsl.nist.gov/incoming). After successful transmission send e-mail to mark.przybocki@nist.gov identifying the name of the file submitted. Alternatively you may send the compressed tar file directly to mark.przybocki@nist.gov.

3.8.2 SYSTEM DESCRIPTION

A valuable tool in discovering strengths and weakness of different algorithmic approaches is the use of system descriptions. Each participant should prepare a brief description of each system submitted and include the description with their submission of results. These system descriptions will be distributed to each participant at the time NIST releases of results.

A typical system description should include:

- A description of the algorithmic approach
- A description of the primary system and all contrastive systems
- A description of the resources required to process the test set, including CPU time and memory

---

[14] To request the evaluation test data, contact Mark Przybocki by e-mail - mark.Przybocki@nist.gov or phone – (301) 975-3347 and indicate when you would like to receive the data.

APPENDIX A – SYSTEM OUTPUT VALUE MODELS

Models of the hypothetic value of ACE system output to hypothetic applications have been created to support the research effort and to provide a basis for evaluation. There are separate (but similar) models for each of the three ACE research tasks. These models are explained in this section.

ENTITY SCORING

The entity evaluation score is defined to be the sum of the values of all system output entities:

$$EDT\_Value_{sys} \quad = \quad \sum_i value\_of\_sys\_entity_i$$

The value of each system output entity is defined to be the product of an inherent entity value and the sum of the values of the entity's mentions:

$$Value_{sys\_entity} \quad = \quad Entity\_Value(sys\_entity) \cdot \sum_m Mention\_Value(sys\_mention_m)$$

The *Entity_Value* of a system output entity is a function of its type. If the output entity is mapped, then the minimum value for the sys entity and its corresponding ref entity is used. For unmapped system entities, *Entity_Value* is weighted by a false alarm penalty. For mapped output entities, *Entity_Value* is discounted for errors in entity type, subtype and class:

$$Entity\_Value \quad = \quad \left\{ \begin{array}{l} \min\begin{pmatrix} ETypeValue(sys) \cdot EClassValue(sys), \\ ETypeValue(ref_{sys}) \cdot EClassValue(ref_{sys}) \end{pmatrix} \cdot \left(W_{Eerr-type} \cdot W_{Eerr-subtype} \cdot W_{Eerr-class}\right) \text{ when mapped} \\ \\ \quad ETypeValue(sys) \cdot EClassValue(sys) \cdot \left(W_{E\text{-}FA}\right) \text{ when entity not mapped} \end{array} \right\}$$

The *Mention_Value* of a system entity mention is also a function of its type. If the mention is mapped, then the minimum value for the sys mention and its corresponding ref mention is used.[15] For mapped system mentions, *Mention_Value* is discounted for errors in mention type, role and style. For unmapped system mentions[16], *Mention_Value* is weighted by a false alarm penalty and a co-reference discount[17]:

$$Mention\_Value \quad = \quad \left\{ \begin{array}{l} \min\begin{pmatrix} MTypeValue(sys), \\ MTypeValue(ref_{sys}) \end{pmatrix} \cdot \left(W_{Merr-type} \cdot W_{Merr-role} \cdot W_{Merr-style}\right) \text{ when mapped} \\ \\ \quad -MTypeValue(sys) \cdot \left(W_{M-FA} \cdot W_{M-CR}\right) \text{ when mention not mapped} \end{array} \right\}$$

For cross-document entities (i.e., for entities that are mentioned in multiple documents), the *Value* of each system entity is accumulated over all documents being evaluated.

---

[15] The mapping of system output mentions to reference mentions is chosen so as to maximize the total value of the mentions.

[16] All mentions of a system output entity are unmapped for entities that are themselves unmapped.

[17] The coreference discount is intended to reduce the penalty for mentions that are valid mentions of an entity but that are incorrectly associated at the entity level. This is considered to be less harmful than mentions that are totally spurious.

RELATION SCORING

The relation evaluation score is defined to be the sum of the values of all system output relations:

$$RDC\_Value_{sys} \quad = \quad \sum_i value\_of\_sys\_relation_i$$

The value of each system output relation is defined to be the product of an inherent relation value and the sum of the values of the relation's entity arguments:

$$Value_{sys\_relation} \quad = \quad Relation\_Value(sys\_relation) \cdot \sum_a Argument\_Value(sys\_argument_a)$$

The *Relation_Value* of a system output relation is a function of its type. If the output relation is mapped, then the minimum value for the sys relation and its corresponding ref relation is used. For unmapped system relations, *Relation_Value* is weighted by a false alarm penalty. For mapped output relations, *Relation_Value* is discounted for errors in relation type and subtype:

$$Relation\_Value \quad = \quad \left\{ \begin{array}{l} \min\begin{pmatrix} RTypeValue(sys), \\ RTypeValue(ref_{sys}) \end{pmatrix} \cdot \left(W_{Rerr-type} \cdot W_{Rerr-subtype}\right) \text{ when mapped} \\ \\ RTypeValue(sys) \cdot \left(W_{R\text{-}FA}\right) \text{ when relation not mapped} \end{array} \right\}$$

The *Argument_Value* of a system relation argument is the *Entity_Value* of that entity argument, where the entity argument of the system relation is mapped to the corresponding argument of the reference relation:[18]

$$Argument\_Value \quad = \quad Entity\_Value(sys)$$

Mapped arguments with an "unacceptably" small *Argument_Value* are assigned an *Argument_Value* of zero.[19]

For cross-document relations (i.e., for relations that are mentioned in multiple documents), the *Value* of each system relation is accumulated over all documents being evaluated. Only those argument entity mentions that appear in these documents are used to compute *Argument_Value*, however.[20]

---

[18] For symmetric relations, argument order is not fixed. In this case, the order used is the order which maximizes the sum of argument values is the order used.

[19] In order for a system output argument to be reasonably considered to represent its corresponding reference argument it is required to exhibit a reasonable overlap with the reference, in terms of *Entity_Value*. Specifically, the *Entity_Value* of the system output argument (mapped to its corresponding reference argument) is compared to the (self-referenced) *Entity_Value* of the corresponding reference argument. A reasonable overlap exists whenever this ratio is greater than or equal to Θ_{Amin}.

[20] The mapping of system arguments to reference arguments is done globally, however, and considers all mentions of the entity arguments. Thus the mapping, while globally optimum, may be suboptimum when considering only a single document.

EVENT SCORING

The event evaluation score is defined to be the sum of the values of all system output events:

$$VDC\_Value_{sys} \quad = \quad \sum_i value\_of\_sys\_event_i$$

The value of each system output event is defined to be the product of an inherent event value and the sum of the values of the event's entity participants:

$$Value_{sys\_event} \quad = \quad Event\_Value(sys\_event) \cdot \sum_p Participant\_Value(sys\_participant_p)$$

The *Event_Value* of a system output event is a function of its type and its modality. If the output event is mapped, then the minimum value for the sys event and its corresponding ref event is used. For unmapped system events, *Event_Value* is weighted by a false alarm penalty. For mapped output events, *Event_Value* is discounted for errors in event type and modality:

$$Event\_Value \quad = \quad \begin{Bmatrix} \min\begin{pmatrix} VTypeValue(sys) \cdot VModeValue(sys), \\ VTypeValue(ref_{sys}) \cdot VModeValue(ref_{sys}) \end{pmatrix} \cdot (W_{Verr-type} \cdot W_{Verr-mode}) \text{ when mapped} \\ \\ VTypeValue(sys) \cdot (W_{V\text{-}FA}) \text{ when event not mapped} \end{Bmatrix}$$

The *Participant_Value* of a system event participant is the *Entity_Value* of that entity participant, where the entity participant of the system event is mapped to the corresponding participant of the reference event.[21] For mapped participants, *Participant_Value* is discounted for errors in participant role. For unmapped system arguments, *Participant_Value* is weighted by a false alarm penalty:

$$Participant\_Value \quad = \quad \begin{Bmatrix} Entity\_Value(sys) \cdot (W_{Perr\text{-}role}) \text{ when mapped} \\ \\ Entity\_Value(sys) \cdot (W_{P-FA}) \text{ when participant not mapped} \end{Bmatrix}$$

Participants with zero *Participant_Value* are considered to be unmapped. Further, mapped participants with an "unacceptably" small *Participant_Value* are assigned a *Participant_Value* of zero.[22]

For cross-document events (i.e., for events that are mentioned in multiple documents), the *Value* of each system event is accumulated over all documents in which the event is mentioned. Only those event entity mentions that appear in these documents are used to compute *Participant_Value*, however.[23]

---

[21] The mapping of the participants of a system output event to those of a reference event is done so as to maximize the sum of the participant values.

[22] In order for a system output participant to be reasonably considered to represent its corresponding reference participant it is required to exhibit a reasonable overlap with the reference, in terms of *Entity_Value*. Specifically, the *Entity_Value* of the system output participant (mapped to its corresponding reference participant) is compared to the (self-referenced) *Entity_Value* of the corresponding reference participant. *A reasonable overlap exists whenever this ratio is greater than or equal to* Θ_{Pmin}.

[23] The mapping of system participants to reference participants is done globally, however, and considers all mentions of the entity arguments. Thus the mapping, while globally optimum, may be suboptimum when considering only a single document.

TIMEX2 SCORING

The timex2 evaluation score is defined to be the sum of the values of all system timex2 outputs:

$$TMX\_Value_{sys} = \sum_i value\_of\_sys\_timex_i$$

The value of each system timex2 output is defined to be the product of two factors – one that accrues value according to the timex2 attributes that are specified, and one which accounts for attribute errors in terms of a product of a discount weights for misrecognized timex2 attributes:

$$Value_{sys\_timex} = Timex\_Value(sys\_timex) \cdot \prod_i W_{Terr-i}(sys\_attribute_i)$$

$$Timex\_Value = \left\{ \begin{array}{ll} \displaystyle\sum_{sys\_attribute} \left\{ \begin{array}{ll} TAttrValue_{sys\_attribute} & \text{if } ref\_attribute \text{ exists} \\ 0 & \text{otherwise} \end{array} \right\} & \text{when mapped} \\ \displaystyle\sum_{sys\_attribute} \left\{ TAttrValue_{sys\_attribute} \right\} & \text{when not mapped} \end{array} \right\}$$

The timex2 attributes are “VAL”, “ANCHOR_DIR”, “ANCHOR_VAL”, “MOD”, and “SET”.

SCORING PARAMETER SETTINGS

The scoring parameters may be adjusted to suit the application. There are currently five sets of parameters available as command line options. In addition to the default parameters there are four other sets, called: "Easy", "Hard", "MaxScore" and "MinScore":

	MinScore	Hard	Default	Easy	MaxScore
Entities:					
ETypeValue					
for **PER**	1.00	1.00	**1.00**	1.00	1.00
for **ORG/VEH/WEA**	0.50	0.50	**0.50**	0.50	0.50
for **GPE**	0.25	0.25	**0.25**	0.25	0.25
for **LOC**	0.10	0.10	**0.10**	0.10	0.10
for **FAC/TMP**	0.05	0.05	**0.05**	0.05	0.05
EClassValue					
for **SPC**	1.00	1.00	**1.00**	1.00	1.00
for **ATR/GEN/NEG/USP**	0.00	0.00	**0.00**	0.00	0.00
MTypeValue					
for **NAM**	1.00	1.00	**1.00**	1.00	1.00
for **NOM/BAR/MWH/PRE**	0.20	0.20	**0.20**	0.20	0.20
for **PRO/HLS/PTV/WHQ**	0.04	0.04	**0.04**	0.04	0.04
for all others	0.00	0.00	**0.00**	0.00	0.00
$W_{Eerr\text{-}type}$	**0.00**	**0.00**	**0.50**	**0.75**	**1.00**
$W_{Eerr\text{-}subtype}$	**0.00**	**0.50**	**0.90**	**1.00**	**1.00**
$W_{Eerr\text{-}class}$	**0.00**	**0.50**	**0.75**	**1.00**	**1.00**
$W_{E\text{-}FA}$	**1.00**	**1.00**	**0.75**	**0.50**	**0.00**
$W_{Merr\text{-}type}$ $W_{Merr\text{-}role}$ $W_{Merr\text{-}style}$	**0.00**	**0.50**	**0.90**	**1.00**	**1.00**
$W_{M\text{-}FA}$	**1.00**	**1.00**	**0.75**	**0.50**	**0.00**
$W_{M\text{-}CR}$	**1.00**	**0.50**	**0.00**	**0.00**	**0.00**
Relations:					
RTypeValue (for all types)	1.00	1.00	**1.00**	1.00	1.00
$W_{Rerr\text{-}type}$	**0.00**	**0.00**	**0.50**	**0.75**	**1.00**
$W_{Rerr\text{-}subtype}$	**0.00**	**0.50**	**0.90**	**1.00**	**1.00**
$W_{R\text{-}FA}$	**1.00**	**1.00**	**0.75**	**0.50**	**0.00**
Θ_{Amin}	**0.75**	**0.50**	**0.25**	**0.00**	**0.00**
Events:					
VTypeValue (for all types)	1.00	1.00	**1.00**	1.00	1.00
VModeValue (for all modalities)	1.00	1.00	**1.00**	1.00	1.00
$W_{Verr\text{-}type}$	**0.00**	**0.00**	**0.50**	**0.75**	**1.00**
$W_{Verr\text{-}mode}$	**0.00**	**0.50**	**0.75**	**1.00**	**1.00**
$W_{V\text{-}FA}$	**1.00**	**1.00**	**0.75**	**0.50**	**0.00**
$W_{Perr\text{-}role}$	**0.00**	**0.00**	**0.50**	**0.75**	**1.00**
$W_{P\text{-}FA}$	**1.00**	**1.00**	**0.75**	**0.50**	**0.00**
Θ_{Pmin}	**0.75**	**0.50**	**0.25**	**0.00**	**0.00**
Timex2s:					
TAttrValue					
for **detection**	0.10	0.10	**0.10**	0.10	0.10
for **ANCHOR_DIR**	0.25	0.25	**0.25**	0.25	0.25
for **ANCHOR_VAL**	0.50	0.50	**0.50**	0.50	0.50
for **MOD**	0.10	0.10	**0.10**	0.10	0.10
for **SET**	0.10	0.10	**0.10**	0.10	0.10
for **VAL**	1.00	1.00	**1.00**	1.00	1.00
$W_{Terr\text{-}ANCHOR\_DIR}$	**0.00**	**0.50**	**0.75**	**0.90**	**1.00**
$W_{Terr\text{-}ANCHOR\_VAL}$	**0.00**	**0.00**	**0.50**	**0.75**	**1.00**
$W_{Terr\text{-}MOD}$	**0.00**	**0.50**	**0.90**	**1.00**	**1.00**
$W_{Terr\text{-}NON\_SPECIFIC}$	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
$W_{Terr\text{-}SET}$	**0.00**	**0.50**	**0.75**	**0.90**	**1.00**
$W_{Terr\text{-}VAL}$	**0.00**	**0.00**	**0.25**	**0.50**	**1.00**
$W_{T\text{-}FA}$	**1.00**	**1.00**	**0.75**	**0.50**	**0.00**

APPENDIX B – THE APF SYSTEM OUTPUT FORMAT

The DTD for the APF XML format is given at the end of this appendix. This DTD accommodates information beyond the minimum required for system evaluation. As a guide to system developers, the minimum information required to support evaluation is listed below.

APF INFORMATION REQUIRED BY THE ACE EVALUATION TOOL

SOURCE FILE INFORMATION – For each source_file element the following are required:

- A URI attribute which specifies the uniform resource identifier of the source data.
- A TYPE attribute which specifies the type of the subsumed documents as one of:
  "text", "audio", or "image".
- A SOURCE attribute which specifies the type of source, typically:
  "newswire", "broadcast news", or "newspaper".
- One or more document elements.

DOCUMENT INFORMATION – For each document element the following are required:

- A DOCID attribute. Each document in the evaluation must have a unique DOCID.
- One or more entity elements.
- Zero or more relation elements.
- Zero or more event elements.

ENTITY INFORMATION – For each entity element the following are required:

- An ID attribute which uniquely identifies the entity. For each document there may be only one entity element per entity. The entity ID must be globally unique. This ID is used to associate cross-document entity information.
- A TYPE attribute which specifies the entity type as one of:
  "PER", "ORG", "VEH", "WEA", "GPE", "LOC", or "FAC".
- An optional SUBTYPE attribute which specifies the entity subtype.
- A CLASS attribute which specifies the entity class as one of:
  "SPC", "GEN", "ATR", "NEG", or "USP".
- One or more mention elements.
- An optional entity_attributes element.

ENTITY MENTION INFORMATION – For each entity_mention element the following are required:

- An optional ID attribute which uniquely identifies the entity mention within the document.
- A TYPE attribute which specifies the mention type as one of:
  "NAM", "NOM", "PRO", or "PRE".
- An optional ROLE attribute which specifies the mention role as one of the entity types.
- An optional specification of style of reference (specifying whether the mention is literal or metonymic). This may be done in three ways, namely using either the attribute STYLE or REFERENCE (with a value of either "LITERAL" or "METONYMIC") or the attribute METONYMY_MENTION (with a value of either "FALSE" or "TRUE").
- An extent element.
- A head element. But if there is no head element, the extent element will be used in its place.

ENTITY NAME INFORMATION – If an entity has an entity_attributes element, entity names are specified using name elements.

RELATION INFORMATION – For each relation element the following are required:

- An ID attribute which uniquely identifies the relation. For each document there may be only one relation element per relation. The relation ID must be globally unique. This ID is used to associate cross-document relation information.
- A TYPE attribute which specifies the relation type as one of:
  "PHYS", "PER-SOC", "EMP-ORG", "ART", "OTHER-AFF", "GPE-AFF", "DISC", or "METONYMY".
- An optional SUBTYPE attribute which specifies the relation subtype.
- Two rel_entity_arg elements.
- Zero or more relation_mention elements.
- Zero or more relation_time elements.

RELATION ARGUMENT INFORMATION – For each rel_entity_arg element the following are required:

- An ARGNUM attribute, namely "1" or "2", which specifies the position of the argument in the relation.
- An ENTITYID attribute which identifies the argument.

RELATION MENTION INFORMATION – For each relation_mention element the following are required:

- An ID attribute which uniquely identifies the relation mention.
- Two rel_mention_arg elements.
- Zero or more rel_mention_time elements.
- An optional ldc_extent element.

RELATION MENTION ARGUMENT INFORMATION – For each rel_mention_arg element the following are required:

- An ARGNUM attribute, namely "1" or "2", which specifies the position of the argument mention in the relation.
- An ENTITYMENTIONID attribute which identifies the argument mention.
- An extent element.

RELATION TIME INFORMATION – For each relation_time element and rel_mention_time element the following are optional:

- A TYPE attribute, which specifies the time type as either "VALUE" or "ANCHOR".
- A VAL attribute.
- A MOD attribute.
- A DIR attribute.
- An optional source element.

EVENT INFORMATION – For each event element the following are required:

- An ID attribute which uniquely identifies the event. For each document there may be only one event element per event. The event ID must be globally unique. This ID is used to associate cross-document event information.

- A TYPE attribute which specifies the event type as one of:
  “BRK’, “MAK”, “GIV”, “MOV”, “INT”, or “OTH”.
- A MODALITY attribute which specifies the event modality as one of:
  “Real” or “NotReal”.
- One or more event_participant elements.
- Zero or more event_mention elements.

EVENT PARTICIPANT INFORMATION – For each event_participant element the following are required:

- An ENTITYID attribute which identifies the participant.
- A ROLE attribute which identifies the participant role as one of:
  “Agent”, “Object”, “Source”, “Target”, “Location”, “Time”, or “Modifier”.

EVENT MENTION INFORMATION – For each event_mention element the following are required:

- An ID attribute which uniquely identifies the event mention.
- A TYPE attribute which specifies the event mention type as one of:
  “NOM” or “SEN”.
- Zero or more event_mention_participant elements.
- An optional extent element.
- An optional anchor element.

EVENT MENTION PARTICIPANT INFORMATION – For each event_mention_participant element the following are required:

- An ENTITYID attribute which uniquely identifies the event mention participant.
- An ENTITYMENTIONID attribute which uniquely identifies the mention of the participant.
- A ROLE attribute which specifies the role of the mention as one of the event participant roles.
- An optional extent element.

TIMEX INFORMATION – For each timex element the following are required:

- An ID attribute which uniquely identifies the timex2. For each document there may be only one timex element per timex2. The timex ID must be globally unique. This ID is used to associate cross-document timex information.
- A TYPE attribute which specifies the event type as “TIMEX2”.
- Zero or more of the following five TIMEX2 attributes:
  “VAL”, “MOD”, “ANCHOR_VAL”, “ANCHOR_DIR”, and “SET”.
- One or more timex_mention elements.
- An optional “COMMENT” attribute.

TIMEX MENTION INFORMATION – For each timex_mention element the following are required:

- An optional ID attribute which uniquely identifies the timex2 within the document.
- Zero or more of the following six TIMEX2 attributes:
  “VAL”, “MOD”, “ANCHOR_VAL”, “ANCHOR_DIR”, “SET”, and “NON_SPECIFIC”.
- An extent element.
- A head element. But if there is no head element, the extent element will be used in its place.

- An optional “COMMENT” attribute.

EXTENT/HEAD/NAME/ANCHOR/SOURCE INFORMATION – Extent/head/name/anchor/score use locator elements to locate data in the source file. Locator elements include charseq, timespan and bblist. These elements are defined in the DTD. The choice of locator elements depends on the TYPE attribute of the source_file.

THE APF DTD

```
<!ELEMENT source_file (document+)>
<!ATTLIST source_file
    URI      CDATA              #REQUIRED
    SOURCE   CDATA              #IMPLIED
    TYPE     (text|audio|image) #REQUIRED
    VERSION  (2.0|3.0|4.0)      #IMPLIED
    AUTHOR   CDATA              #IMPLIED
    ENCODING CDATA              #IMPLIED
>

<!ELEMENT document (entity+,relation*,event*)>
<!ATTLIST document
    DOCID    CDATA              #REQUIRED
>

<!-- ********************************************** -->
<!-- Entities, their attributes and their mentions. -->
<!-- ********************************************** -->

<!ELEMENT entity (entity_mention+,entity_attributes?)>
<!ATTLIST entity
    ID      CDATA                             #REQUIRED
    TYPE    (PER|ORG|LOC|GPE|FAC|VEH|WEA|TMP) #REQUIRED
    SUBTYPE (Government|Commercial|Educational|
             Non-Profit|Other|Address|Boundary|Celestial|
             Water-Body|Land-Region-Natural|Region-Local|
             Region-Subnational|Region-National|
             Region-International|Continent|Nation|
             State-or-Province|County-or-District|
             Population-Center|Building|Subarea-Building|
             Bounded-Area|Conduit|Path|Barrier|Plant|Land|
             Air|Water|Subarea-Vehicle|Blunt|Exploding|Sharp|
             Chemical|Biological|Shooting|Projectile|
             Nuclear)                         #IMPLIED
    CLASS   (NEG|SPC|GEN|USP)                 #REQUIRED

>

<!ELEMENT entity_mention (extent,head?)>
<!ATTLIST entity_mention
    ID              CDATA                                #REQUIRED
    TYPE            (NAM|NOM|PRO|PRE)                    #REQUIRED
    LDCTYPE         (NAM|NOM|BAR|MWH|PRO|WHQ|PRE|HLS|MSC|
                     PTV|CMC|APP|ARC|DE|PCN|PMM|EPM|EAP) #IMPLIED
    ROLE            (PER|ORG|LOC|GPE|FAC)                #IMPLIED
    STYLE           (LITERAL|METONYMIC)                  #IMPLIED
    REFERENCE       (LITERAL|METONYMIC)                  #IMPLIED
    METONYMY_MENTION (TRUE|FALSE)                         #IMPLIED
    LDCATR           (TRUE|FALSE)                         #IMPLIED
>

<!ELEMENT entity_attributes (name*)>

<!ELEMENT name       (bblist|charspan|charseq|timespan)>
<!ELEMENT extent     (bblist|charspan|charseq|timespan)>
<!ELEMENT head       (bblist|charspan|charseq|timespan)>
<!ELEMENT anchor     (bblist|charspan|charseq|timespan)>
<!ELEMENT source     (bblist|charspan|charseq|timespan)>
<!ELEMENT ldc_extent (bblist|charspan|charseq|timespan)>

<!ELEMENT charspan (#PCDATA)>
<!ATTLIST charspan
    START NMTOKEN  #REQUIRED
    END   NMTOKEN  #REQUIRED
>

<!ELEMENT charseq (#PCDATA)>
<!ATTLIST charseq
    START NMTOKEN  #REQUIRED
    END   NMTOKEN  #REQUIRED
>

<!ELEMENT timespan         (#PCDATA)>
<!ATTLIST timespan
    START NMTOKEN  #REQUIRED
    END   NMTOKEN  #REQUIRED
>

<!ELEMENT bblist (boundingbox+)>

<!ELEMENT boundingbox (#PCDATA)>
<!ATTLIST boundingbox
    PAGE NMTOKEN   #REQUIRED
    X    NMTOKEN   #REQUIRED
    DX   NMTOKEN   #REQUIRED
    Y    NMTOKEN   #REQUIRED
    DY   NMTOKEN   #REQUIRED

<!-- ********************************************** -->
<!-- RELATIONS, their attributes and their mentions. -->
<!-- ********************************************** -->

<!ELEMENT relation (rel_entity_arg, relation_mention*,
                    relation_time*)>
<!ATTLIST relation
    ID       CDATA                     #REQUIRED
    TYPE     (PHYS|PER-SOC|EMP-ORG|ART|OTHER-AFF|
              GPE-AFF|DISC|METONYMY)   #REQUIRED
    SUBTYPE  (Located|Near|Part-Whole|Business|
              Family|Other|Employ-Executive|
              Employ-Staff|Employ-Undetermined|
              Member-of-Group|Subsidiary|Partner|
              User-or-Owner|Inventor-or-Manufacturer|
              Ethnic|Ideology|Citizen-or-Resident|
              Based-In)                #IMPLIED
>

<!ELEMENT rel_entity_arg (#PCDATA)>
<!ATTLIST rel_entity_arg
    ENTITYID CDATA  #REQUIRED
    ARGNUM   CDATA  #REQUIRED
>

<!ELEMENT relation_mention (ldc_extent?, rel_mention_arg+,
                            rel_mention_time*)>
<!ATTLIST relation_mention
    ID                   CDATA             #REQUIRED
    LDCLEXICALCONDITION  (Possessive|Preposition|
                          PreMod|Formulaic|Verbal|
                          Participial)     #REQUIRED
>

<!ELEMENT rel_mention_arg (extent?)>
<!ATTLIST rel_mention_arg
    ENTITYMENTIONID CDATA  #REQUIRED
    ARGNUM          CDATA  #REQUIRED
>

<!ELEMENT relation_mention_time (source*)>
<!ATTLIST relation_mention_time
    TYPE  (VALUE|ANCHOR) #REQUIRED
    VAL   CDATA          #REQUIRED
    MOD   CDATA          #IMPLIED
    DIR   CDATA          #IMPLIED
>

<!-- ************************************************ -->
<!-- LDC: EVENTS, their attributes and their mentions. -->
<!-- ************************************************ -->

<!ELEMENT event (event_participant+, event_mention*)>
<!ATTLIST event
    ID       CDATA                       #REQUIRED
    TYPE     (BRK|MAK|GIV|MOV|INT|OTH)  #REQUIRED
    MODALITY (Real|NotReal)              #REQUIRED
```

```
>

<!ELEMENT event_participant (extent?)>
<!ATTLIST event_participant
    ENTITYID        CDATA   #REQUIRED
    ROLE   (Agent|Object|Source|Target|
            Location|Time|
            Modifier)       #REQUIRED
>

<!ELEMENT event_mention (extent, anchor,
                         event_mention_participant*)>
<!ATTLIST event_mention
    ID   CDATA      #REQUIRED
    TYPE (NOM|SEN)  #REQUIRED
>

<!ELEMENT event_mention_participant (extent?)>
<!ATTLIST event_mention_participant
    ENTITYID        CDATA   #REQUIRED
    ENTITYMENTIONID CDATA   #REQUIRED
    ROLE   (Agent|Object|Source|Target|
            Location|Time|
            Modifier)       #REQUIRED
    SOURCE (Event|Relation) #REQUIRED
>

<!-- ******************************************** -->
<!-- TIMEX2S, their attributes and their mentions. -->
<!-- ******************************************** -->

<!ELEMENT timex (timex_mention+)>
<!ATTLIST event
    ID            CDATA        #REQUIRED
    TYPE          (TIMEX2)     #REQUIRED
>

<!ELEMENT timex_mention (head?, extent?)>
<!ATTLIST timex_mention
    ID            CDATA        #IMPLIED
    TYPE          (TIMEX2)     #IMPLIED
    VAL           CDATA        #IMPLIED
    MOD           CDATA        #IMPLIED
    ANCHOR_DIR    CDATA        #IMPLIED
    ANCHOR_VAL    CDATA        #IMPLIED
    SET           CDATA        #IMPLIED
    NON_SPECIFIC  CDATA        #IMPLIED
    COMMENT       CDATA        #IMPLIED
>
```

APPENDIX C – NIST CTM FORMAT (ASR OUTPUT)

As described in section 3.3, the ACE English Broadcast News evaluation data will be distributed in two formats:

- Ground truth transcripts (manually created)
- Speec-To-Text (STT) transcripts (generated by an automatic speech recognizer engine with an estimated word error rate of 7-8%)

CTM FORMAT

CTM SOURCE FILE INFORMATION – The STT source files will be tailored to include the necessary document tags for ACE. The actual STT output will be identified with "<W" tags (sword tag). The following is a partial example:

```
<DOC>
<DOCNO> CNN20001211.1400.0303 </DOCNO>
<DOCTYPE SOURCE=”broadcast news”> NEWS STORY </DOCTYPE>
<BODY>
<DATE_TIME> 12/11/2000 14:09:32.25 </DATE_TIME>
<TEXT>
<W recid=NA Bsec=303.64 Dur=0.16 Clust=NA Conf=NA> HELLO
<W recid=NA Bsec=303.80 Dur=0.26 Clust=NA Conf=NA> FROM
<W recid=NA Bsec=304.06 Dur=0.33 Clust=NA Conf=NA> WASHINGTON
<W recid=NA Bsec=303.39 Dur=0.30 Clust=NA Conf=NA> STATE
<W recid=NA Bsec=303.69 Dur=0.36 Clust=NA Conf=NA> UNIVERSITY
<W recid=NA Bsec=304.05 Dur=0.31 Clust=NA Conf=NA> IN
…
</TEXT>
</BODY>
</DOC>
```